AD662069

(1)

Technical Research Note 192

AD

RELATIONSHIP OF EXPRESSED CONFIDENCE TO ACCURACY OF TRANSCRIPTION BY OPERATIONAL COMMUNICATIONS PERSONNEL

by Eugene P. Stichman

Combat Systems Research Division

OCTOBER 1967

U. S. Army
Behavioral Science Research Laboratory

30

BLANK PAGES
IN THIS
DOCUMENT
WERE NOT
FILMED

BEHAVIORAL SCIENCE RESEARCH LABORATORY

An activity of the Chief, Research and Development

J. E. UHLANER
Director

M. O. BECKER
Colonel, GS
Commanding

NOTICES

DDC AVAILABILITY: Qualified requestors may obtain copies of this report directly from DDC. Available, for sale to the public, from the Clearinghouse for Federal Scientific and Technical Information, Department of Commerce, Springfield, Virginia 22151.

DISTRIBUTION: Primary distribution of this report has been made by BESRL. Please address correspondence concerning distribution of reports to: U. S. Army Behavioral Science Research Laboratory, Washington, D. C. 20315. Copies may also be obtained on loan from local depository libraries. A list of these libraries may be obtained from: Documents Expediting Project, Library of Congress, Washington, D. C. 20540.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U. S. Army Behavioral Science Research Laboratory.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

AD

Technical Research Note 192

RELATIONSHIP OF EXPRESSED CONFIDENCE TO ACCURACY OF TRANSCRIPTION BY OPERATIONAL COMMUNICATIONS PERSONNEL

by Eugene P. Stichman

COMBAT SYSTEMS RESEARCH DIVISION
Philip J. Bersh, Chief

U. S. ARMY BEHAVIORAL SCIENCE RESEARCH LABORATORY

Office, Chief Research and Development
Department of the Army

Washington, D. C. 20315

October 1967

Army Project Number 2J024701A723

Combat Communications b-22

This document has been approved for public release and sale; its distribution is unlimited.

BESRL Technical Research Reports and Technical Research Notes are intended for sponsors of R&D tasks and other research and military agencies. Any findings ready for implementation at the time of publication are presented in the latter part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

The COMBAT COMMUNICATIONS Task employs controlled laboratory experimentation in studies designed to improve the overall performance of personnel involved in tactical communications operations. Concentrating for the present on voice communications, the research seeks to attain greater speed, accuracy, and completeness in the extraction of information from voice-radio and telephone media. Three primary objectives are: (1) to increase the efficiency of radio-telephone communications in a tactical environment; (2) to enhance the performance of transcribers and analysts in the extraction of information from communications media; and (3) to develop improved human factors techniques for tactical electronic countermeasures.

A previous study (TRN 175) dealt with the ability of personnel untrained in communications to rate their own performance in receiving and transcribing voice-radio messages embedded in noise. The present study sought to determine whether operational communications personnel could rate their performance with greater precision.

The research was conducted under Subtask b, "Development of improved work methods for message transmission, reception, and transcription", FY 1967 Work Program. In addition to research on confidence ratings, studies are conducted to improve the operator's performance through such factors as redundancy, repetition, enhanced discrimination of speech sounds, and additional transcription methods.

J. E. UHLANER, Director
U. S. Army Behavioral Science
Research Laboratory

RELATIONSHIP OF EXPRESSED CONFIDENCE TO ACCURACY OF TRANSCRIPTION BY OPERATIONAL COMMUNICATIONS PERSONNEL

BRIEF

Requirement:

To determine whether experienced communications operators are able to rate their performance in transcribing voice radio messages partially embedded in noise with sufficient precision for the ratings to have potential operational utility.

Procedure:

Eight experienced communications operators rated their confidence in the accuracy of their reception and transcription of messages received at three signal-to-noise ratios (-6 db, 0 db, +6 db). A five-point rating scale was used. As a control, they also transcribed messages without making confidence ratings. Measures of transcript accuracy and expressed confidence in transcription obtained under the experimental conditions were compared with results from a prior study in which the subjects were neither communications operators nor trained in any communications procedures prior to experimental familiarization.

Findings:

The experienced communications operators were highly successful in judging the accuracy of their transcription, achieving a close relationship between confidence rating and performance (r_{tet} = .78), although overconfidence at the upper end of the scale and underconfidence at the lower end were evident.

Intelligibility improved from 20% to 88% as signal-to-noise ratio increased.

The experienced communications operators performed better than the non-communications trained subjects in the former study both in accuracy of transcription and in precision of confidence ratings. In neither study was average accuracy of the transcripts affected by having subjects judge their transcription.

In both studies, subjects tended to make effective use of less than all five points of the confidence rating scale.

Utilization of Findings:

The practicability of obtaining operationally useful expressions of confidence from transcribers was strongly supported, although the most effective form for a standardized confidence rating procedure remains to be determined. Standardized ratings could assist communications analysts and decision makers, permitting them to weight the transcribed information appropriately and to place it in proper perspective with respect to data from other sources.

RELATIONSHIP OF EXPRESSED CONFIDENCE TO ACCURACY OF TRANSCRIPTION BY OPERATIONAL COMMUNICATIONS PERSONNEL

CONTENTS

	Page
INTRODUCTION	1
METHOD	1
Experimental Design	2
Subjects	2
Stimulus Material	2
Apparatus	2
Work Methods	3
Confidence Rating	3
Procedure	4
RESULTS	4
CONCLUSIONS	10
LITERATURE CITED	12
DISTRIBUTION	13
APPENDIX	15
DD Form 1473 (Document Control Data - R&D)	27

FIGURES

Page

Figure 1. Regression of accuracy on confidence across signal-to-noise ratios 5

2. Regression of accuracy on confidence by subject across signal-to-noise ratios 6

3. Mean intelligibility scores as a function of signal-to-noise ratio 7

4. Regression of accuracy on confidence for each signal-to-noise ratio 8

5. Comparison of the regressions of confidence on accuracy for both samples with the ideal regression 9

RELATIONSHIP OF EXPRESSED CONFIDENCE TO ACCURACY OF TRANSCRIPTION BY OPERATIONAL COMMUNICATIONS PERSONNEL

Magnetic tape recording of incoming messages is standard procedure in many different voice radio telephone communications operations. The recordings are used in a variety of ways, including re-transmittal in radio relay operations and transcription into hard copy for subsequent analysis in decision making operations. When a message is partially masked by noise, it is very difficult for the operator to receive and transcribe the entire message correctly. Unless communications are being jammed, the unwanted noise tends to be sporadic, and the intelligibility of different sections of the message varies inversely with the amount of unwanted noise. The communications transcriber often has subjective impressions of confidence about the accuracy with which he is able to transcribe such partially masked messages.

Preliminary research, using personnel without formal training or experience in communications, has shown a positive relationship between the transcriber's confidence in his correct reception and his accuracy of transcription (1). While far from ideal for operational use, this relationship was sufficient to warrant further research using operational communications personnel. The existence of a close relationship between confidence ratings of performance and accuracy of transcription among experienced operators would be of considerable value in the development of improved standing operating procedures. The improved procedures could be applied to all communications operations where information must be transmitted, extracted, and assimilated. Reliable measures of transcriber ability to relate confidence to accuracy also could provide the communications analyst with important time-saving clues. Such measures could afford objective estimates of the necessity for additional transcriptions of a message received under marginal or less than marginal listening conditions (2). More important, by establishing differential levels of acceptance for sections of transcripts on the basis of the transcriber's confidence judgments, the analyst might be able to extract more reliable information from the transcript of a partially masked transmission.

The present study dealt with the ability of operational communicators to evaluate their own performance in extracting information from noise-embedded voice radio communications.

METHOD

In an operational communications situation, the operators, monitors, and transcribers rarely know the listening conditions under which they must operate from moment to moment or from message to message. Measurement of performance under different signal-to-noise ratios was therefore necessary to obtain information about behavior across listening conditions. In the present study, measures of two aspects of performance--transcription accuracy and expressed confidence in the correctness of the transcription--were obtained at each of three signal-to-noise ratios representing a broad range of listening conditions. These measures were analyzed to determine the relationship between the confidence rating and transcription accuracy.

Experimental Design

The design was a 3 x 2 x 8 factorial, three signal-to-noise ratios constituting the first factor, two work methods the second factor, and eight enlisted men the third factor. In the first work method, the transcribers assigned confidence ratings to each transcribed word. In the second work method, no confidence ratings were made. Each man performed under all six combinations of factors one and two.

Subjects

The subjects were eight enlisted men selected at random from a population of school trained, highly experienced operational communicators. All eight were in PULHES hearing category 1 (supported by MAICO Model H-1 Audiometer hearing tests).[1] All men had had some field experience in the required MOS and also experience in transcription.

Stimulus Material

The stimulus material consisted of the 1,000 phonetically-balanced monosyllabic words developed by the Harvard Psycho-Acoustics Laboratory (3). These 1,000 words are divided into 20 lists each consisting of 50 words. Five complete randomizations of the 20 lists, prerecorded on tape, were used. The words in each list were presented at an intensity of approximately 75 decibels (0.0002 dynes per cm^2), one word every four and one-half seconds, at signal-to-noise ratios of +6 db, 0 db, and -6 db. Each word was preceded by the carrier sentence: "YOU WILL TRA ___(word)___."

Apparatus [1]

Word lists were reproduced on an Ampex tape recorder (Model 351) and electronically mixed (Ampex MX-35 Mixer) with noise from a Bruel and Kjaer Random Noise Generator (No. 1402). The mixed output was amplified (Macintosh MC-75) and presented binaurally through headphones (Telex, 600 ohm). A double-walled audiometric research sound booth was used both for training and for data collection.

---

[1] Identification of instruments and materials is included solely for precision in reporting experimental procedures and does not constitute indorsement of any commercial product by the Department of the Army.

Work Methods

Subjects listened to and transcribed word lists under each of three signal-to-noise ratios, rating their confidence in the correctness of each word as they transcribed it. They also listened to and transcribed the same lists under the same signal-to-noise ratios without making any expressions of confidence. Transcription of a word list while making the confidence ratings was the experimental condition; transcription of the list without making the confidence rating was the control condition. Order of presentation of the two conditions was randomized to control for possible order effects.

Confidence Rating [2]

Five categories of expressed confidence were used:

5 I AM FULLY CONFIDENT THAT I RECEIVED THE WORD CORRECTLY.

4 I AM SUBSTANTIALLY CONFIDENT THAT I RECEIVED THE WORD CORRECTLY.

3 I AM MODERATELY CONFIDENT THAT I RECEIVED THE WORD CORRECTLY.

2 I AM SLIGHTLY CONFIDENT THAT I RECEIVED THE WORD CORRECTLY.

1 I AM NOT AT ALL CONFIDENT THAT I RECEIVED THE WORD CORRECTLY.

Ratings would be completely accurate if all words rated 5 were correctly transcribed, all words rated 1 were incorrect, and half of all words rated 3 were correct, with about three-fourths of all words rated 4 and one-fourth of all words rated 2 correct. Subjects were instructed to apply the following concept in making their ratings: A rating of 5 was to be assigned when the subject would bet a large sum that his reception and transcription of a word was in fact correct. Conversely, he was to assign a rating of 1 to a word when he would not think at all of betting on its correctness. He was to assign a rating of 3 when he felt that the word was one of two he could have chosen, and ratings of either 2 or 4 when he felt that his confidence fell midway between categories 3 and 1 and categories 5 and 3, respectively.

---

[2] A more complete discussion of the rational for this rating procedure may be found in the report of the earlier research (1).

Procedure

Subjects were trained in four groups of two men each[3]. Training was in accordance with established procedures for speech intelligibility testing (4). One and one-half days of familiarization, using two of the five randomizations of the word lists, brought all subjects to approximately equal level of familiarity with the stimuli, with the general transcription procedures required for the experiment, and with the three signal-to-noise ratios. An additional half day of training was devoted to familiarization with the confidence rating scale and the experimental conditions. Each pair of subjects was then tested for twelve experimental sessions, spread over three days, using the remaining three randomizations of the word lists as stimuli. Each session consisted of listening to and transcribing 10 word lists. The experimental sessions were 50 minutes in length. Each pair of subjects had a rest period of approximately one-half hour between experimental sessions, with a one-hour lunch break after the first two sessions each day.

The measure of intelligibility was the mean percentage of words correctly transcribed for all lists. This measure was obtained both for each of the three signal-to-noise ratios and for combined performance across signal-to-noise ratios. The measure was obtained separately under experimental and control conditions.

The measure of accuracy obtained under the experimental condition was the percentage of words given any one rating which were transcribed correctly.

RESULTS

There was a relatively high relationship between the confidence which subjects expressed in the correctness of their transcripts and the accuracy of received messages. Measured across subjects and the three signal-to-noise ratios, the coefficient of correlation[4] between confidence and accuracy was +.78. At each of the five confidence rating steps, mean accuracy scores were significantly different from each other ($p < .001$), and mean accuracy scores increased in linear fashion with the confidence rating. Summaries of these analyses are presented in Tables A-1 and A-2

---

[3] Due to duty assignments, subjects were available only in pairs and only for one consecutive five-day period. Familiarization time was therefore shortened, and the number of experimental sessions per day was doubled as compared with the earlier research (1).

[4] All computed correlation coefficients were tetrachoric. This measure was obtained by collapsing the 2 x 5 (right--wrong x rating scale) distribution into a 2 x 2 (right--wrong x high--low ratings) distribution, splitting the rating array as near the median as possible.

of the Appendix; Table A-3 provides the overall mean accuracy of words rated at each step on the confidence rating scale. The slope of the linear regression of accuracy on confidence across signal-to-noise ratios was .184. The slope of the hypothetical ideal linear regression of accuracy on confidence would be .25. The regression function and the mean accuracy at each step of the confidence rating scale are shown in Figure 1 for both observed performance and ideal performance. To achieve the ideal, mean accuracy scores would have to be 0%, 25%, 50%, 75%, and 100%, for confidence ratings 1 through 5, respectively.

Overconfidence at the upper end of the function and underconfidence at the lower end were observed (subjects rated incorrectly transcribed words high and rated correctly transcribed words low). Eighteen percent of the words rated 5 by all subjects were incorrectly transcribed, and ten percent of the words rated 1 were received and transcribed correctly.

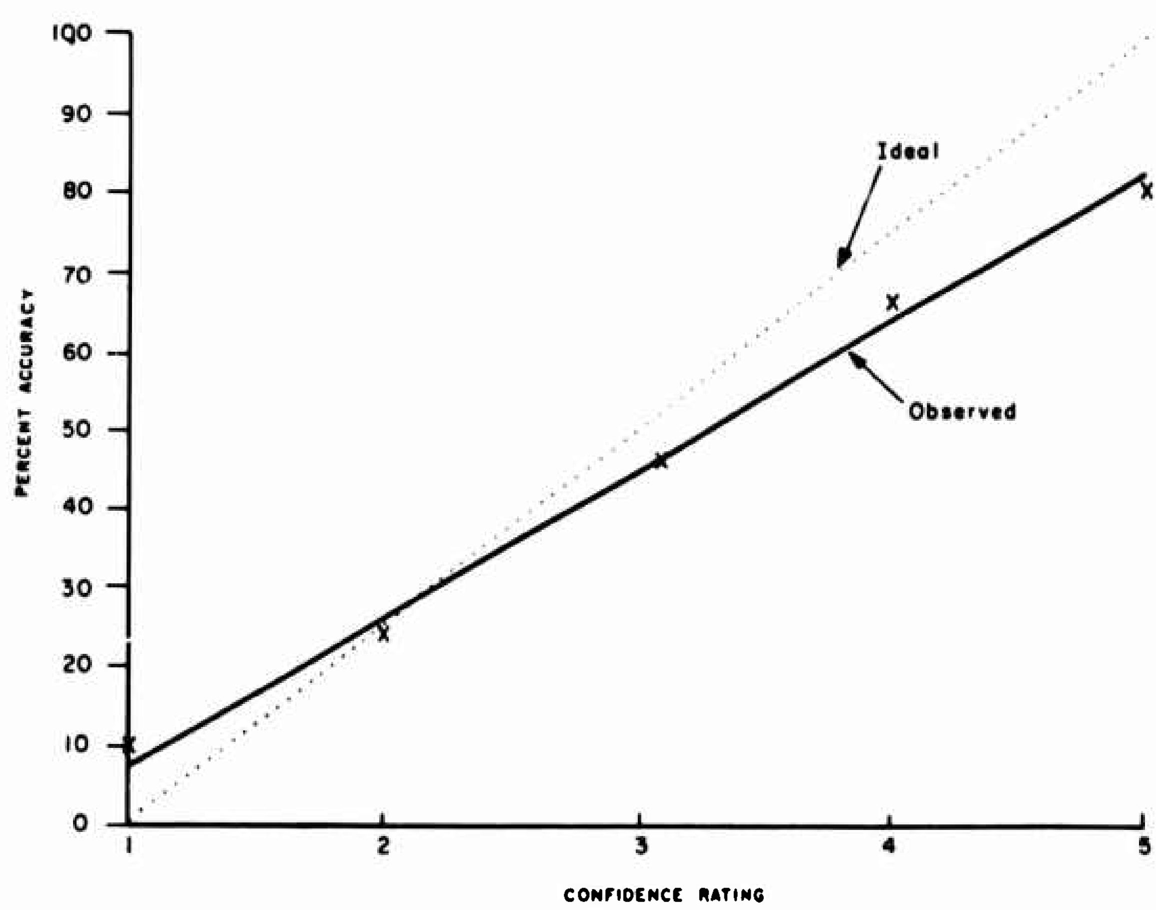
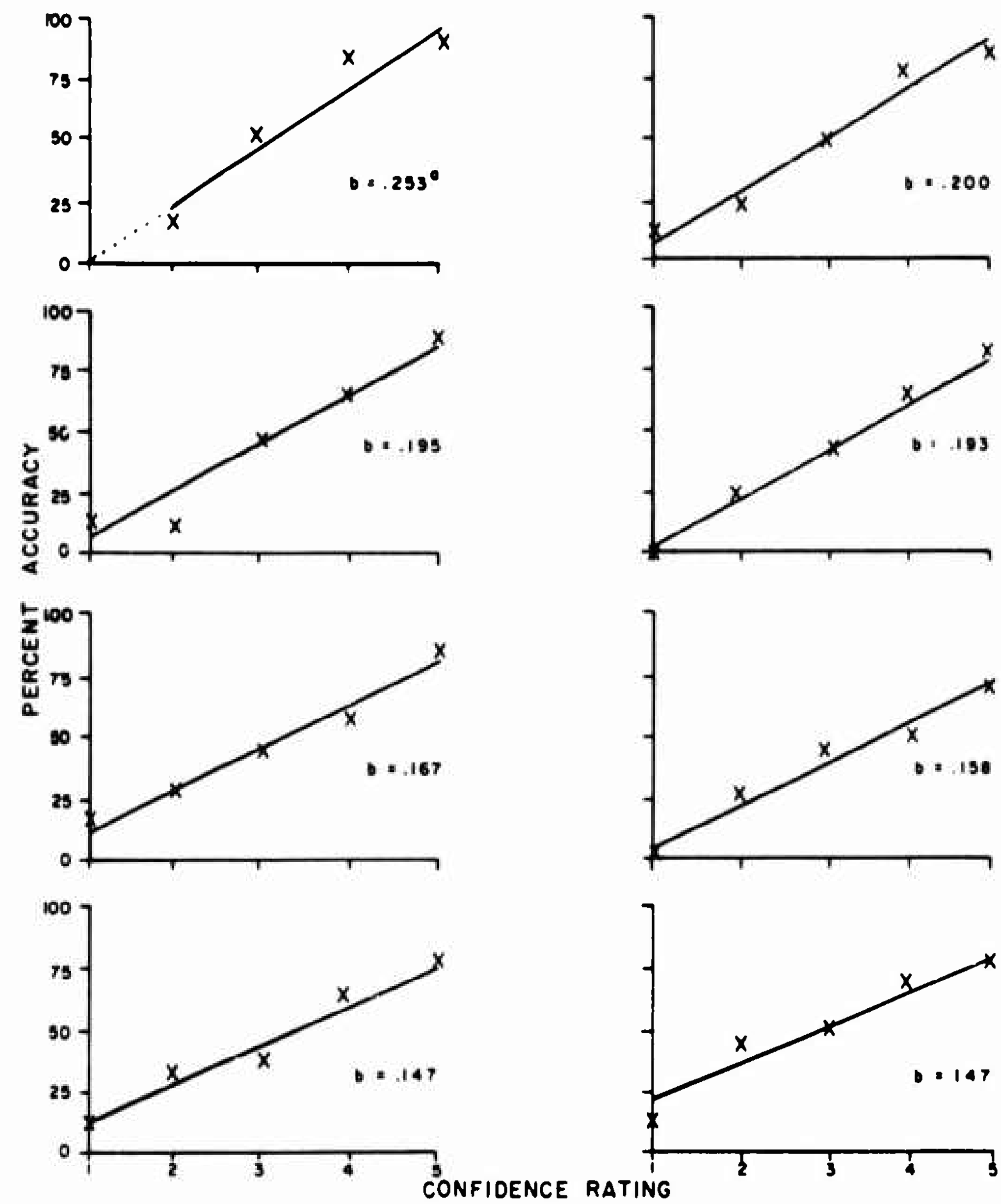


Figure 1. Regression of accuracy on confidence across signal—to—noise ratios

Although individual differences were observed among subjects, every subject showed a close relationship between confidence and accuracy. For each subject, mean accuracy scores at each of the five confidence rating steps were significantly different from each other ($p < .01$), and these means increased in linear fashion with the confidence rating. The eight graphs in Figure 2 present both mean accuracy scores at each confidence rating and regression functions separately for each subject across signal-to-noise ratios. Where mean accuracy scores at adjacent confidence ratings for some subjects seem very close, significance was nonetheless obtained because of the substantial number of determinations at some rating steps.



[a] Slope was projected from four points. Subject made no correct responses rated 1.

FIGURE 2. Regression of accuracy on confidence by subject across signal—to—noise ratios

As expected, intelligibility improved as a direct function of the signal-to-noise ratio, increasing from a mean of approximately 20 percent to approximately 88 percent. Figure 3 compares means obtained at each signal-to-noise ratio under both experimental (rated) and control (non-rated) conditions. Effect of signal-to-noise ratio on these intelligibility means was significantly different from chance ($p < .001$). Having subjects assign confidence ratings did not significantly affect mean intelligibility. Table A-4 presents the means and standard deviations of the intelligibility scores, and Table A-5 shows the summary of the analysis of variance.

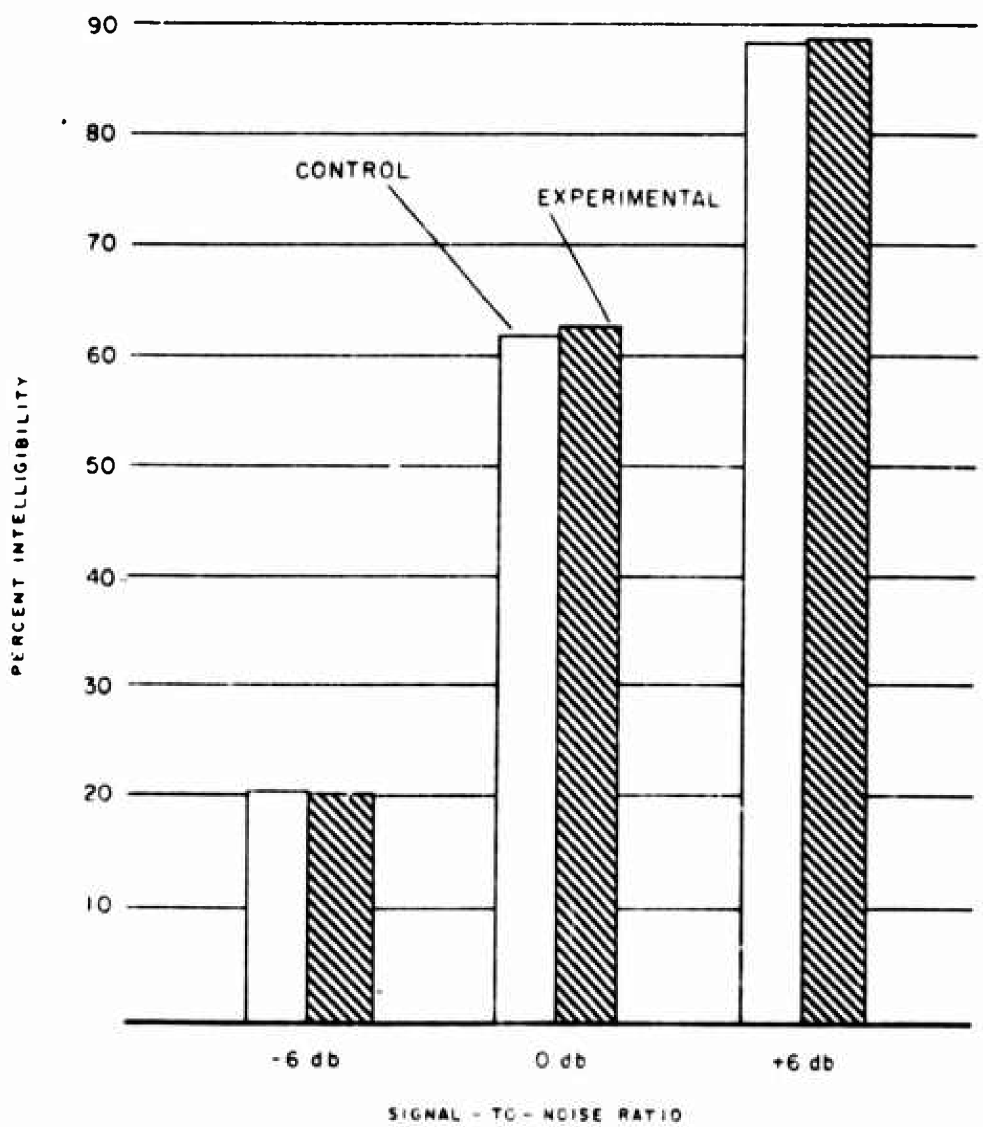


Figure 3. Mean intelligibility scores as a function of signal—to—noise ratio

Mean confidence ratings also increased as a direct function of signal-to-noise ratio. Mean ratings at the different signal-to-noise ratios were significantly different from each other ($p < .001$). Mean confidence ratings and significance of difference values for these means are given in Table A-6.

Since both mean intelligibility and mean confidence were significantly affected by signal-to-noise ratio, the results for the three signal-to-noise ratios were analyzed separately. Correlation coefficients[5] between confidence and accuracy were +.49, +.48, and +.53 for the -6 db, the 0 db, and the +6 db signal-to-noise ratios, respectively. At each signal-to-noise ratio, mean accuracy scores for the five confidence rating steps were significantly different from each other ($p < .001$). Moreover, at each signal-to-noise ratio, mean accuracy increased in a substantially linear fashion with confidence rating, although a slight curvature (the quadratic component) was apparent at the -6 db signal-to-noise ratio[6]. Summaries of these analyses are presented in Table A-7.

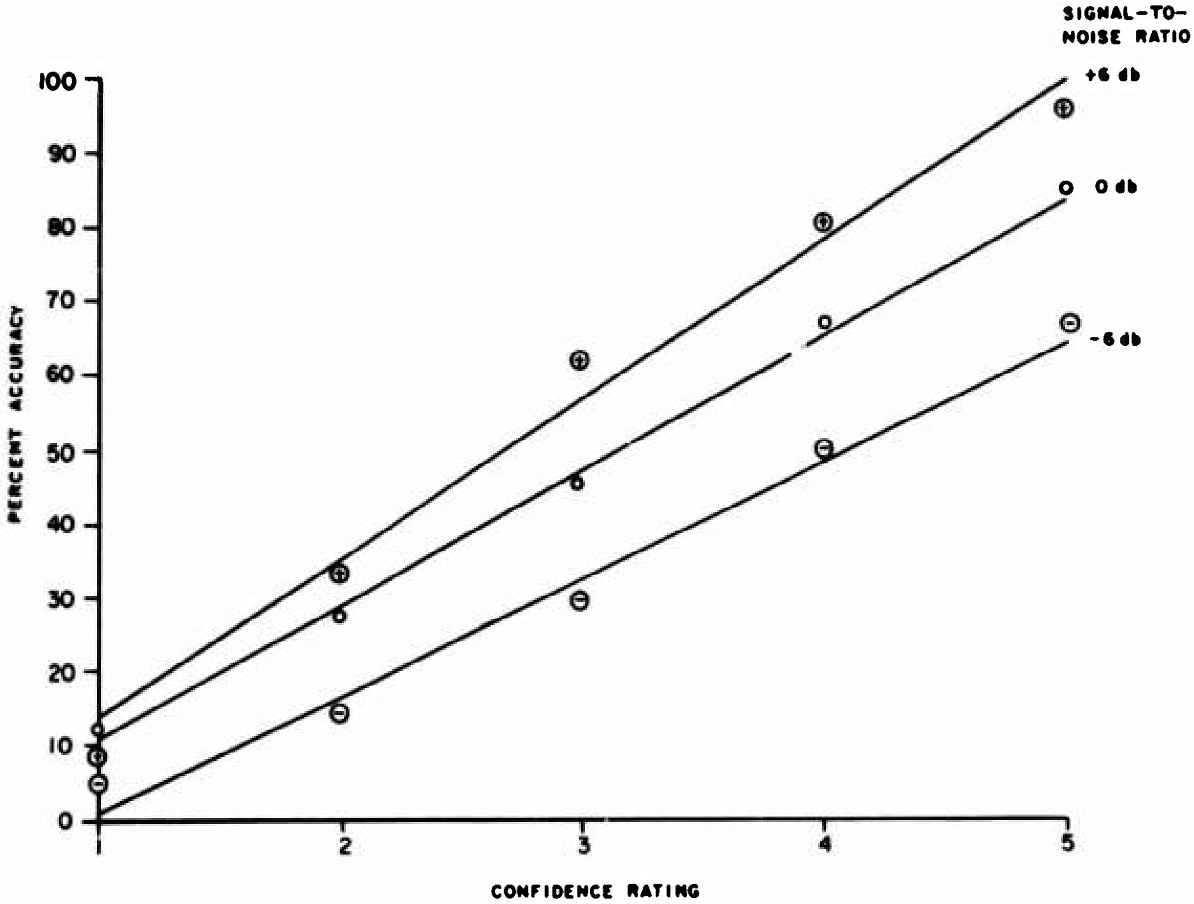


Figure 4. Regression of accuracy on confidence for each signal to-noise-ratio

---

[5] See footnote 4.

[6] This deviation form linearity, while not significant ($.10 < p < .05$), was caused by underconfidence at the -6 db listening condition.

Separate plots of accuracy as a function of confidence at each of the three signal-to-noise ratios are shown in Figure 4. The slopes of the linear regressions describing performance at each of the three signal-to-noise ratios were .156 at -6 db, .179 at 0 db, and .211 at +6 db. The slopes of the regression of accuracy on confidence were significantly affected by the signal-to-noise ratio ($p < .01$)[7]. That this significant interaction is itself linear can be seen from Table A-2.

None of the subjects in this study made effective use of all five steps in the confidence rating scale (Table A-8). The large number of high ratings was a result of the relatively high intelligibility at both the 0 db and the +6 db signal-to-noise ratios.

The sample of operational communications personnel in the present study clearly outperformed[8] the sample of enlisted men in the earlier research (1) who had had no previous formal training or experience in

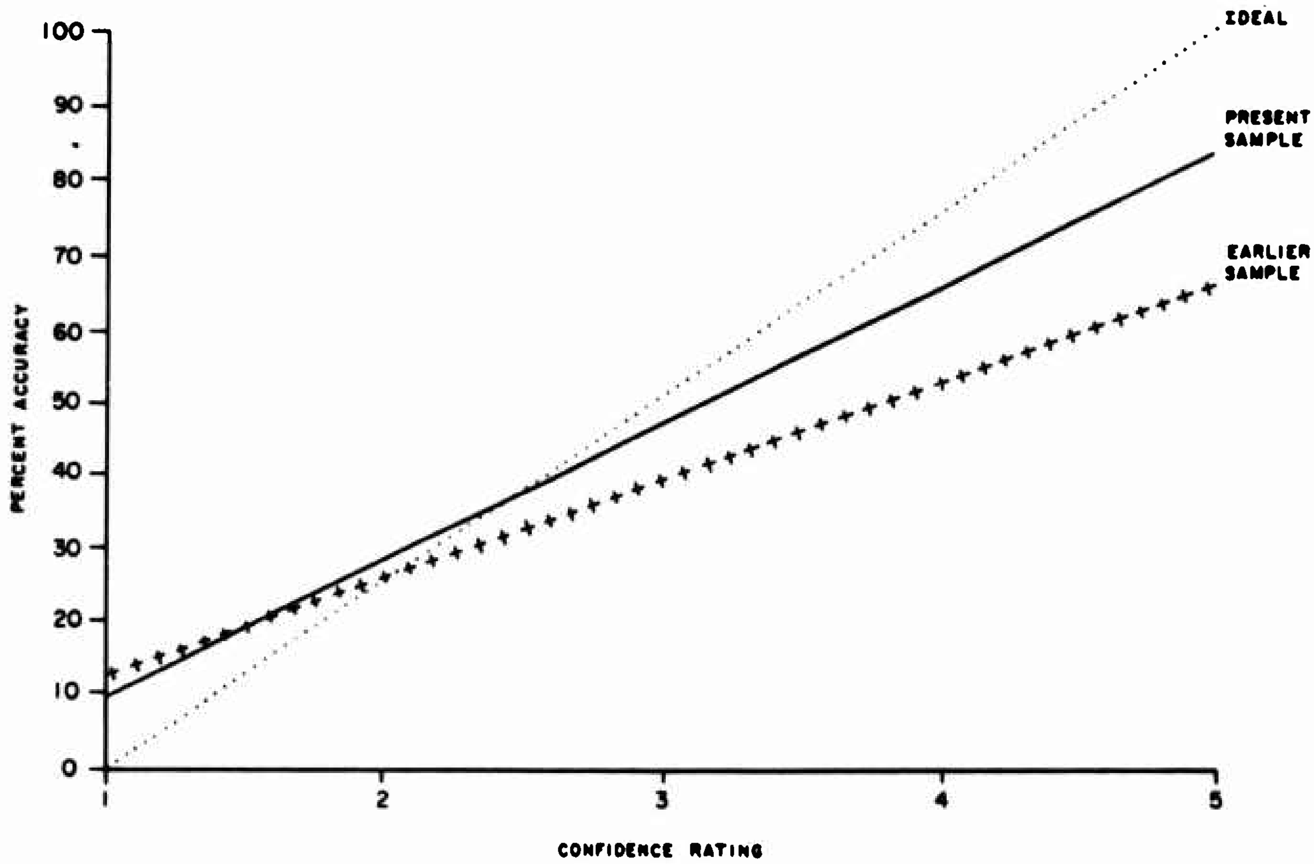


Figure 5. Comparison of the regressions of confidence on accuracy for both samples with the ideal regression

[7] Homogeneity of regression (5).

[8] The results were compared across listening conditions. The unpredictability of moment-to-moment noise interference with voice radio communications in the field makes the relationship between confidence and accuracy averaged over all listening conditions the best practical basis for prediction.

either communications procedures or transcription techniques. From the analysis summary in Table A-9, performance means were significantly different from each other ($p < .01$) and the regression of confidence on accuracy was significant ($p < .001$). Although the two samples exhibited similar performance trends, a much closer relationship between confidence and transcription accuracy was shown by the operational sample in the present study (correlation coefficient of .78 as opposed to .57). While overconfidence at the upper end of the rating scale and underconfidence at the lower end were observed in both studies, the magnitude of the observed deviations from the ideal in the present sample was considerably smaller than in the earlier sample. In the present study, fewer than 18% of the responses rated 5 were incorrect (compared with 32% in the earlier study), and only 10% of the responses rated 1 were correct (compared with 13%). Differences between the two samples in the relationship between confidence and accuracy become even more apparent when performance is compared with an ideal where overconfidence and underconfidence are both non-existent (Figure 5).

CONCLUSIONS

In spite of the procedural differences which favored subjects in the earlier research--longer familiarization period and fewer sessions per day--the sample of operational communications personnel in the present study outperformed the sample in the earlier study. Their formal school training in general communications procedures coupled with field experience in voice-radio message transcription under degraded conditions evidently enabled the operational communicators to "read through noise" and transcribe more accurately. While overconfidence and underconfidence still occurred, the magnitude of such errors was less. Ability to rate one's own performance on the job would appear to be directly related to experience. It is likely that the well-trained communications operator implicitly performs some type of evaluating while he is transcribing, drawing on his past experience to do so. The present study provides strong indication that trained operators can provide operationally useful confidence ratings without having their performance affected by the act of rating. Formulation of a standardized rating scale therefore becomes practicable.

Had the additional familiarization time and the three additional testing days been available in the present study, as in the earlier research, the operational communicators might have even more closely approximated the ideal in rating their performance. The decrease in effectiveness of the confidence rating as a function of the degradation of the message might also have been reduced with additional training. This is especially important because, even under less-than-marginal listening conditions, the rating measure affords a valuable basis for differential weighting of the rated portions of a message. It can be seen from Figure 4 that mean accuracy at the -6 db signal-to-noise ratio varied from approximately 6% at the lower end to approximately

67% at the upper end of the rating scale, yet intelligibility at this signal-to-noise ratio was only 20%. The introduction of some type of standardized rating scale in the MOS course training could therefore prove helpful.

While some question might be raised regarding the potential deterioration of performance as a result of the extra work required in rating each message segment as it is transcribed, the data from both the earlier research (1) and the present study argue strongly against this possibility. In no case was the performance using ratings significantly different from its control (see Figure 3 and Table A-4 of this study and the corresponding figure and table from the earlier report).

Table A-8 and the corresponding table in the earlier report reveal that the majority of subjects utilized only three levels of confidence--high, medium, and low--in their ratings. These three levels of confidence do not correspond to any of the actual points on the rating scale itself, although the inference is easy to make. The actual ratings on the scale which were effectively used varied among the subjects. Only one or two subjects used four scale points effectively. Insufficient familiarization and training in the use of the ratings, less than adequate instructions regarding them, or the short time interval between message presentations (three seconds from the end of one to the onset of the next) may have been primary causes. For any or all of these reasons, the five-point rating scale simply may not be the best type to use in transcription evaluation of this nature. If a standardized rating procedure is to be introduced into the MOS course or implemented in the field, the significant determinants of rating effectiveness must be conclusively identified.

Ultimately, the value of the confidence rating procedure for implementation depends on the minimization of errors of overconfidence and underconfidence. The results of the present study, in comparison with those obtained with the earlier sample, suggest that introducing adequate instruction in assigning confidence ratings as part of the formal MOS school curriculum would improve the relationship between confidence and accuracy by reducing overconfidence and underconfidence, and provide a basis for the meaningful differential weighting of severely degraded messages. Implementation of reliable transcriber confidence judgments would provide communications analysts and decision-makers with an objective and workable measure of the accuracy of transcripts of degraded messages. The procedure would assist the analyst by placing transcribed information in the proper perspective and allowing the decision-makers to weigh this information properly with respect to data from other sources. The overall result would be a more efficient and more reliable extraction of information from noise-embedded voice radio-telephone communications.

LITERATURE CITED

1. Stichman, E. P. Transcriber confidence in relation to accuracy of transcription. BESRL Technical Research Note 175. July 1966.

2. Stichman, E. P. and G. E. Renaud. Information extraction from voice communications: Work methods for single transcribers. BESRL Technical Research Note 154. June 1965.

3. Egan, J.P. Articulation testing methods. Laryngscope, 59, 1948, pp 955-991.

4. American standard method for measurement of monosyllabic word intelligibility. American Standards Association. No. S3.2-1960. New York: May 1960.

5. Winer, B. J. Statistical principles in experimental design. McGraw-Hill. New York: 1962.

SELECTED BIBLIOGRAPHY

Castelnovo, A. E., J. G. Tiedemann, and D. A. Dobbins. Performance of single vs multiple voice radio transcribers working under three speech to noise ratios. BESRL Technical Research Note 135. September 1963.

Castelnovo, A. E., J. G. Tiedemann, and D. M. Skordahl. Individual differences in transcribing voice radio messages embedded in atmospheric noise. BESRL Technical Research Note 137. October 1963.

DISTRIBUTION

U. S. Army Behavioral Science Research Laboratory

DISTRIBUTION LIST

Directorate for Armed Forces I and E
Director, Army Research, OCRD
Director, Army Research, DD & Scientific Dir. (2)
Deputy Chief of Staff for Personnel
Assistant Chief of Staff for Force Development (3)
Assistant Chief of Staff for Intelligence
Chief of Personnel Operations, DA (5)
CG, U. S. Continental Army Command (22)
CG, U. S. Army Combat Development Command (2)
CO, U. S. Army Enlisted Evaluation Center (2)
Chief of Information, DA
Chief of Chaplains, DA
Assistant Secretary of Defense for Education
CG, Automatic Data Field Systems Command
Comdt., Marine Corps
Director, Human Resources Research Office (2)
Directors of Research, HumRRO Field Divisions
U. S. Army Medical Research Laboratory, Psychology Division
CO and Director, U. S. Naval Training Devices Center
CG, U. S. Army CEDEC
CG, U. S. Army Electronic Proving Ground
OIC, U. S. Naval Medical NP Research Unit
Director, WRAIR, Walter Reed Army Medical Center
Chief, Personnel Research Staff, OP, U. S. Department of Agriculture
USES, BES, U. S. Department of Labor
The Adjutant General's Office, Personnel Services Support Directorate
Chief of Naval Personnel (4)
Office of Naval Research (8)
CRESS, Information Sys. Br.
Director, National Security Office
Director, Central Intelligence Agency
Chief, Office of Personnel, PHS, Department of Health, Education, and Welfare
Office of the Provost Marshall General
Office of the Surgeon General, DA
CG, U. S. Army Materiel Command
CG, U. S. Army Security Agency
Director of Rsch and Dev., U. S. Army Electronics Command
Head, Psychology Labs., U. S. Army Natick Laboratories
CG, Aberdeen Proving Ground, Hum Engr Labs
Chief, Bu M and S, Department of the Navy
Director, U. S. Naval Research Laboratory

CO, U. S. Army Garrison Tech. Ref. Office (2)
Office, Director of WAC
CINPAOAF (IGSAG)
U. S. Army MED R&D Command, Biomed. Stress Rsch. Br.
USAERDL - Human Factors Engr Tech & Rsch Spt Dept
Director, USA Engr Rsch and Dev Labs., Fort Belvoir
CO. U. S. Army Research Office (Durham)
CO, U. S. Army R and D Group (FE)
Chief, U. S. Army R and D Liaison Group (Eur)
Chief, U. S. Army R and D Office (Alaska)
Chief, U. S. Army R and D Office (Panama)
Chief, U. S. Army R and D Office, U. S. Army Arctic Test Center
CO, U. S. Army Edgewood Arsenal
Cmdr, Tactical Air Recon Cntr, Shaw AFB
Chief, Intelligence Materiel Dev Office (EWL), Fort Holabird
CG, Air R and D Command
CO, Hq. U. S. Air Force (AFNICAD)
CO, Hq. U. S. Air Force (AFPDPCE)
Chief, Officer Rsch and Review Br, U. S. Coast Guard Hq
U. S. Army Standardization Group (Canada)
U. S. Army Standardization Group (UK)
Comdt., Command and General Staff College
Director, Military Psychology and Leadership, USMA (2)
Director of Research, Office of Research, USMA
Superintendent, U. S. Air Force Academy (3)
Director of Admission, U. S. Coast Guard Academy
Cmdt., U. S. Army Management School (2)
CG, U. S. Army Infantry Center
Cmdt., U. S. Army Artillery and Missle School
Cmdt., U. S. Army Missle and Munitions Center and School
Cmdt., USAF Air Ground Operations School
Cmdt., U. S. Army Engineer School
Cmdt., USAF School of Aerospace Med. (2)
U. S. Army Air Defense Board and School, Fort Bliss (3)
U. S. Army Aviation Test Board and School Library, Fort Rucker (2)
Director of Instruction, U. S. Army Special Warfare School
Educational Advisor, U. S. Coast Guard Training Center
Cmdt., U. S. Army War College, Carlisle Barracks
Cmdt., USASA Training Center and School
Director of Instruction, U. S. Army Armor School
Superintendent, U. S. Naval PG School
Dean, Marine Corps Institute
U. S. Army Infantry Board
Library of Congress, Exchange and Gift Division (6)
Army Library
Library of Congress, Unit X, Documents Expediting Project (105)
Defense Documentation Center (20)

APPENDIX

Tables of Results of Data on Transcriber Confidence and Transcription Accuracy

		Page
Table A-1.	Summary of accuracy score analysis of variance	17
A-2.	Trend Analysis	18
A-3.	Mean accuracy at each confidence rating across signal-to-noise ratios	19
A-4.	Intelligibility means and standard deviations by work method and signal-to-noise ratio	19
A-5.	Summary of intelligibility score analysis of variance	20
A-6.	Mean confidence rating and "t" values by signal-to-noise ratio	20
A-7.	Summary of accuracy score analysis of variance and trend analysis by signal-to-noise ratio	21
A-8.	Percentage of words assigned each rating by subject across signal-to-noise ratios	24
A-9.	Summary of accuracy score analysis of variance comparing trained and untrained communicator samples	25

Table A-1

SUMMARY OF ACCURACY SCORE ANALYSIS OF VARIANCE

Source	SS	DF	MS	F
Between:				
Subjects (A)	0.2715	7	0.0388	
Within:				
Confidence (B)	7.5606	4	1.8902	102.173[a]
B by A	0.5169	28	0.0185	
Signal-to-noise Ratio (C)	1.2562	2	0.6281	35.891[b]
C by A	0.2445	14	0.0175	
BC	0.1244	8	0.0156	2.364
BC by A	0.3579	54[c]	0.0066	
TOTAL	10.3320	117		

[a]F (4,28) .001 = 6.25

[b]F (2,14) .001 = 11.78

[c]Two cells had no entries.

Table A-2

TREND ANALYSIS

(1) OVERALL TREND ACROSS SIGNAL-TO-NOISE RATIOS

Source	SS	DF	MS	F
Linear Component	7.5331	1	7.5331	1141.38[a]
Quadratic Component	0.0044	1	0.0044	N S
Deviations	0.0231	2	0.0116	N S
Error	0.3579	54	0.0066	

[a]$F\ (1,54)\ .001 = 12.16$

(2) DIFFERENCES COMPARING SIGNAL-TO-NOISE RATIOS

Source	SS	DF	MS	F
Linear Component	0.0715	2	0.0358	5.42[a]
Quadratic Component	0.0410	2	0.0205	N S
Deviations	0.0119	4	0.0060	N S
Error	0.3579	54	0.0066	

[a]$F\ (2,54)\ .01 = 5.04$

Table A-3

MEAN ACCURACY AT EACH CONFIDENCE RATING ACROSS SIGNAL-TO-NOISE RATIOS

	Rating				
	1	2	3	4	5
Mean Accuracy (percent)	10	26	46	66	82

Table A-4

INTELLIGIBILITY MEANS AND STANDARD DEVIATIONS BY
WORK METHOD AND SIGNAL-TO-NOISE RATIO

Signal-to-Noise Ratio	Work Method			
	Control		Experimental	
	Mean	SD	Mean	SD
+6 db	88.26	6.52	88.37	6.54
0 db	62.76	11.64	63.34	10.37
-6 db	20.52	7.90	20.19	7.80

Table A-5

SUMMARY OF INTELLIGIBILITY SCORE ANALYSIS OF VARIANCE

Source	DF	MS	F
Subjects (S)	7	275.85	
Work Method (W)	1	3.15	N S
W by S	7	8.11	
Signal-to-Noise Ratio (R)	2	377,577.85	9,933.645[a]
R by S	14	38.01	
WR	2	16.66	N S
WR by S	14	18.60	
TOTAL	47		

[a] $F\ (2,14)\ .001 = 11.78$

Table A-6

MEAN CONFIDENCE RATING AND "t" VALUES
BY SIGNAL-TO-NOISE RATIO

	Signal-to-Noise Ratio		
	1	2	3
Mean Confidence	2.19	3.88	4.58
σ_m	0.035	0.029	0.022
"t" values			
1	——	10.43*	16.21*
2		——	5.48*
3			——

$*p < .001$

Table A-7

SUMMARY OF ACCURACY SCORE ANALYSIS OF VARIANCE BY SIGNAL-TO-NOISE RATIO

(1) -6 db SIGNAL-TO-NOISE RATIO

(A) Analysis of Variance

Source	SS	DF	MS	F
Between:				
Subjects (S)	0.2319	7	0.0331	
Within:				
Confidence (C)	1.9653	4	0.4913	49.63[a]
S by C	0.2777	28	0.0099	
TOTAL	2.4749	39		

[a] $F(4,28)$ .001 = 6.25

(B) Trend Analysis

Source	SS	DF	MS	F
Linear Component	1.9251	1	1.9251	194.45[a]
Quadratic Component	0.0360	1	0.0360	3.64[b]
Deviations	0.0042	2	0.0021	NS
Error	0.2777	28	0.0099	

[a] $F(1,28)$ .001 = 13.50

[b] $F(1,28)$ .10 = 2.89

Table A-7
(continued)

(2) 0 db SIGNAL-TO-NOISE RATIO

(A) Analysis of Variance

Source	SS	DF	MS	F
Between:				
Subjects (S)	0.0948	7	0.0135	
Within:				
Confidence ($\dot{C}$)	2.5721	4	0.6430	126.08[a]
C by S	0.1438	28	0.0051	
TOTAL	2.8107	39		

[a] $F(4,28)$ .001 = 6.25

(B) Trend Analysis

Source	SS	DF	MS	F
Linear Component	2.5668	1	2.5668	503.29[a]
Quadratic Component	0.0004	1	0.0004	NS
Deviations	0.0049	2	0.0024	NS
Error	0.1438	28	0.0051	

[a] $F(1,28)$ .001 = 13.50

Table A-7
(continued)

(3) +6 db SIGNAL-TO-NOISE RATIO

(A) Analysis of Variance

Source	SS	DF	MS	F
Between:				
Subjects (S)	0.1894	7	0.0270	
Within:				
Confidence (C)	3.1477	4	0.7869	33.34[a]
C by S	0.6598	28	0.0236	
TOTAL	3.9969	39		

[a]F(4,28) .001 = 6.25

(B) Trend Analysis

Source	SS	DF	MS	F
Linear Component	3.1126	1	3.1126	131.89[a]
Quadratic Component	0.0089	1	0.0089	NS
Deviations	0.0262	2	0.0131	NS
Error	0.6598	28	0.0236	

[a]F(1,28) .001 = 13.50

Table A-8

PERCENTAGE OF WORDS ASSIGNED EACH RATING BY SUBJECT ACROSS SIGNAL-TO-NOISE RATIOS

	Rating				
Subjects	1	2	3	4	5
1	24.29	11.73	12.33	13.53	38.13
2	18.50	5.47	12.63	22.80	40.60
3	13.60	10.10	23.33	9.03	43.93
4	15.97	9.57	18.20	15.00	41.26
5	1.67	19.00	35.63	25.20	18.58
6	5.43	23.60	14.70	29.67	26.60
7	11.17	14.63	14.87	5.83	53.50
8	9.73	12.87	18.70	14.03	44.67
$\bar{X}$	12.54	13.37	18.80	16.89	38.40

Table A-9

SUMMARY OF ACCURACY SCORE ANALYSIS OF VARIANCE COMPARING TRAINED AND UNTRAINED COMMUNICATOR SAMPLES

(1) ANALYSIS OF VARIANCE

Source	SS	DF	MS	F
Between:				
Groups (G)	0.1037	1	0.1037	9.971[a]
Error (a)	0.1460	14	0.0104	
Within:				
Confidence (C)	3.9582	4	0.9900	126.923[b]
C by G	0.1206	4	0.0302	3.872[c]
Error (b)	0.4360	56	0.0078	
TOTAL	4.7645	79		

[a] $F_{(1,14)}$ .01 = 8.86
[b] $F_{(4,40)}$ .001 = 5.70
[c] $F_{(4,40)}$ .01 = 3.83

(2) INTERACTION TREND ANALYSIS

Source	SS	DF	MS	F
Linear Component	0.1025	1	0.1025	13.141[a]
Deviations	0.0181	3	0.0060	N S
Error (b)	0.4360	56	0.0078	

[a] $F_{(1,40)}$ .001 = 12.61

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) U. S. Army Behavioral Science Research Laboratory, OCRD	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

RELATIONSHIP OF EXPRESSED CONFIDENCE TO ACCURACY OF TRANSCRIPTION BY OPERATIONAL COMMUNICATIONS PERSONNEL

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)

Eugene P. Stichman

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
October 1967	34	5

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. DA R&D PJ No. 2J024701A723	Technical Research Note 192
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

Qualified requestors may obtain copies of this report directly from DDC. Available, for sale to the public, from the Clearinghouse for Federal Scientific and Technical Information, Department of Commerce, Springfield, Virginia 22151.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	U. S. Army Security Agency Office, Chief of Rsch and Dev, DA

13. ABSTRACT

The study sought to determine whether operational communications personnel can rate their performance in transcribing voice radio messages partially embedded in noise with sufficient precision for the ratings to have potential operational utility. Eight experienced communications operators rated their confidence in the accuracy of their reception and transcription of messages received at three signal-to-noise ratios (-6 db, 0 db, +6 db), using a five-point rating scale. As a control, subjects also received and transcribed the messages without making ratings. Measures of transcript accuracy and expressed confidence in transcription were compared with results from a prior study in which subjects had no formal training or experience in communications or transcription (Technical Research Note 175). Experienced operators were highly successful in judging their own accuracy, the relationship between confidence and accuracy being $r_{tet} = .78$. Some overconfidence at the upper end of the rating scale and underconfidence at the lower end were evident. Intelligibility improved from 20% to 88% as signal-to-noise ratio increased. The communications operators performed better than the non-communications trained subjects in the former study both in accuracy of transcription and in precision of confidence ratings. Judging the transcription did not affect the average accuracy of the transcripts in either study. In both studies, subjects tended to make effective use of less than all five points of the confidence rating scale.

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Communications research Singal-to-noise ratio Intelligibility scores Accuracy scores Voice-radio communications--transcription procedure Communications monitoring Communications performance measurement Laboratory facilities Confidence rating procedure Information extraction						